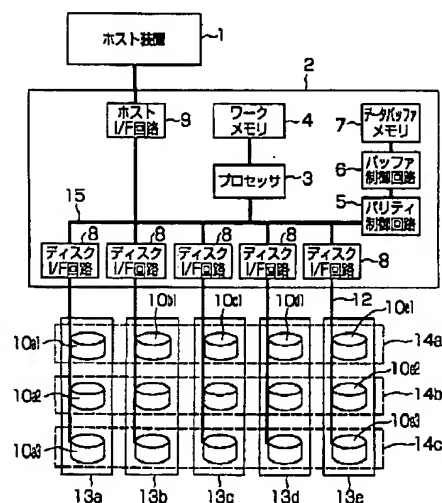


(11)特許出願公開番号



【特許請求の範囲】

【請求項1】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記各ディスクドライブをグループ間の列毎にまとめて前記ディスクコントローラに対して着脱自在にユニット化してなることを特徴とするディスクドライブシステム。

【請求項2】 請求項1記載のディスクドライブシステムにおいて、前記ディスクコントローラは、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に前記複数の中の所定ユニットをスペアユニットとして運用することを特徴とするディスクドライブシステム。

【請求項3】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、

前記ディスクコントローラは、前記複数の中のある一つのディスクドライブが故障したとき、前記グループ毎に前記パリティデータのデータを基に他のディスクドライブのデータのパリティチェックを行い、故障ディスクドライブのデータを復元するデータ復元手段と、

前記データ復元手段により復元されたデータと、前記故障ディスクドライブの属するユニット内の他のディスクドライブのデータとを前記複数の中の所定ユニットのそれぞれに対応するディスクドライブに書き込むデータ書込手段と、

前記データ書込手段によりデータの書き込まれた所定ユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備したことを特徴とするディスクドライブシステム。

【請求項4】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、そのグループ単位にデータ処理を実行するディスクコントローラとを設けてなるディスクドライブシステムにおいて、

前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、

に、

前記ディスクコントローラは、前記複数のユニットの中の一つをパリティユニットとし、そのパリティユニットに他のユニットのディスクドライブが故障したときのパリティチェック用のパリティデータを記録するパリティデータ記録手段と、あるユニット内でディスクドライブが故障したときに、前記パリティユニットの対応するディスクドライブから前記パリティデータを読み出し、それを基に故障ディスクドライブのデータを復元するデータ復元手段と、前記複数のユニットの中の一つをスペアユニットとして、前記データ復元手段により復元された故障ディスクドライブのデータを前記スペアユニットの対応するディスクドライブに書き込む第1の書込手段と、前記故障ディスクドライブが属するユニット内の他のディスクドライブのデータを前記スペアユニットのそれぞれに対応するディスクドライブに書き込む第2の書込手段と、

前記各データの書き込まれた前記スペアユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備したことを特徴とするディスクドライブシステム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、例えば複数台のディスクドライブを制御するRAID方式などを採用したディスクシステムに関する。

【0002】

【従来の技術】従来から、RAID方式のディスクシステム、例えばディスクアレイサブシステムなどでは、あるディスクドライブが故障したときの交換作業性を考慮して筐体設計をしており、このため筐体形状の制約が多いばかりか、筐体内のディスク実装密度も低く改善が望まれている。

【0003】図4に示すように、従来のディスクアレイサブシステムは、個々のディスクドライブ41a～41eの故障を想定してディスクドライブ1台づつを交換する単位として各ユニット42a～42e内に收容し、それをキャビネット40の表面あるいは裏面に沿って一列（アレイ状）に実装するようにしている。

【0004】この場合、システムが稼働中に、あるユニットのディスクドライブが故障しても、キャビネット40の表面側の扉を開ければ全てのユニット42a～42eが表れ、故障ディスクドライブのユニットを容易に取り外し、新しいものに交換することができる。

【0005】ところで、通常、この種のディスクアレイサブシステムを設計する場合、そのキャビネットの縦、横および奥行きなどの寸法は、安定性などを得る面から予めある程度の寸法に決められている。したがって、上記配置条件でキャビネット40内にディスクドライブの

ユニット42a～42eを配置した後、コントローラや電源装置などのユニット44を配置すると、その内部スペースに空間45ができることになる。

【0006】

【発明が解決しようとする課題】しかしながら、これでは、キャビネット内部の利用効率が悪いばかりか、実装できるディスクドライブ数が非常に少ないため、ディスクドライブの実装効率も悪いという問題がある。

【0007】そこで、ディスクドライブの交換性のある程度無視して、キャビネットの奥行き方向に複数のディスクドライブを配置し、それをコントローラのインタフェースに接続することが考えられるが、この場合、どのディスクドライブが故障してもよいようにキャビネットの周囲を扉だらけにするしかなく、またシステム稼働中（他のディスクドライブが動作しているときなど）にインタフェースから故障ディスクドライブを取り外すと、そのときにノイズが発生し、同一インタフェースに接続されている他のディスクドライブにノイズが入り込み、信号が不安定な状態になるため、通常、適当なノイズ対策（ハード的またはソフト的な対策手段）を付加する必要がある、単にキャビネットの空きスペースにディスクドライブを増設するわけにはいかないという問題があった。

【0008】本発明はこのような課題を解決するためになされたもので、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することのできるディスクシステムを提供することを目的としている。

【0009】

【課題を解決するための手段】上記した目的を達成するために、請求項1記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記各ディスクドライブをグループ間の列毎にまとめて前記ディスクコントローラに対して着脱自在にユニット化してなることを特徴としている。

【0010】また請求項2記載のディスクドライブシステムは、請求項1記載のディスクドライブシステムにおいて、前記ディスクコントローラは、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に前記複数の中の所定ユニットをスペアユニットとして運用することを特徴としている。

【0011】さらに請求項3記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復

元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、前記ディスクコントローラは、前記複数の中のある一つのディスクドライブが故障したとき、前記グループ毎に前記パリティディスクのパリティデータを基に他のディスクドライブのデータの復元するデータ復元手段と、前記データ復元手段により復元されたデータと、前記故障ディスクドライブの属するユニット内の他のディスクドライブのデータとを前記複数の中の所定ユニットのそれぞれに対応するディスクドライブに書き込むデータ書込手段と、前記データ書込手段によりデータの書き込まれた所定ユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備している。

【0012】また請求項4記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、そのグループ単位にデータ処理を実行するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、前記ディスクコントローラは、前記複数のユニットの中の一つをパリティユニットとし、そのパリティユニットに他のユニットのディスクドライブが故障したときのパリティチェック用のパリティデータを記録するパリティデータ記録手段と、あるユニット内でディスクドライブが故障したときに、前記パリティユニットの対応するディスクドライブから前記パリティデータを読み出し、それを基に故障ディスクドライブのデータを復元するデータ復元手段と、前記複数のユニットの中の一つをスペアユニットとして、前記データ復元手段により復元された故障ディスクドライブのデータを前記スペアユニットの対応するディスクドライブに書き込む第1の書込手段と、前記故障ディスクドライブが属するユニット内の他のディスクドライブのデータを前記スペアユニットのそれぞれに対応するディスクドライブに書き込む第2の書込手段と、前記各データの書き込まれた前記スペアユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備している。

【0013】

【作用】請求項1記載の発明では、複数のディスクドライブが、所定数毎にグループ化されたものが列設され、各ディスクドライブはディスクコントローラに対してグループ間の列毎にまとめて着脱自在にユニット化されている。

【0014】したがって、あるユニット内のディスクドライブが故障した場合に、そのディスクドライブのデータを復元した後、そのディスクドライブが属するユニットの全てのデータを他のユニットへ複写し、故障したディスクドライブの属するユニットをディスクコントローラから取り外せば、ノイズが発生してもユニットごとに取り外すので、同一インターフェイスに接続された他のディスクドライブも取り外されるため、これによる性能の劣化は起こらなくなる。

【0015】また請求項2記載の発明では、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に複数の中の所定ユニットをスペアユニットとして運用するので、不具合が発生したユニットを抜去しても、システム全体の性能が低下することがなくなる。

【0016】さらに請求項3記載の発明では、複数の中のある一つのディスクドライブが故障したとき、ディスクコントローラのデータ復元手段により、グループ毎にパリティディスクのパリティデータを基に他のディスクドライブのデータのパリティチェックが行われて、故障ディスクドライブのデータが復元される。

【0017】そしてこの復元されたデータと、故障ディスクドライブの属するユニット内の他のディスクドライブのデータとがデータ書込手段により複数の中の所定ユニットのそれぞれ対応するディスクドライブに書き込まれる。データが書き込まれると、所定ユニットは故障ディスクドライブの属するユニットの代替えとして制御手段により運用される。

【0018】すなわち、あるユニットで故障が起これば、その内容は全て所定ユニットに再現されて、その機能も全て所定ユニットによって代替される。

【0019】したがって、故障ディスクドライブの属するユニットごと交換しても、性能劣化がなくなる。

【0020】また請求項4記載の発明では、複数の中のある一つのディスクドライブが故障したとき、ディスクコントローラのデータ復元手段により、グループ毎にパリティディスクのパリティデータを基に他のディスクドライブのデータのパリティチェックが行われて、故障ディスクドライブのデータが復元される。

【0021】すると、まず、復元されたデータが、第1の書込手段によりスペアユニットの対応するディスクドライブに書き込まれる。続いて、第2の書込手段により、故障ディスクドライブが属するユニット内の他のディスクドライブのデータがスペアユニットのそれぞれ対応するディスクドライブに書き込まれる。

【0022】そして各データの書き込まれたスペアユニットは、制御手段により故障ディスクドライブの属するユニットの代替えとしてシステム上で運用される。

【0023】したがって、故障ディスクドライブの属するユニットごと交換しても、性能劣化が起こらなくなる。

る。

【0024】上記により、故障ディスクドライブを新しいディスクドライブに交換するためにシステムからそのユニットを取り外してもノイズの影響がなくなり、システム動作中でも他のユニットに影響を与えることなく交換することができる。

【0025】この結果、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することができる。

【0026】

【実施例】以下、本発明の実施例を図面を参照して詳細に説明する。

【0027】図1は本発明に係る一実施例のディスクアレイサブシステムの構成を示す図である。

【0028】同図において、1はホスト装置であり、例えば計算機などである。2はディスクコントローラであり、RAID機能を備えたディスクアレイの制御装置である。3はプロセッサであり、ディスクコントローラ2の機能を制御するものである。4はワークメモリであり、プロセッサ3により利用されるメモリである。5はパリティ制御回路であり、RAID機能に基づきパリティデータを生成すると共に、そのパリティデータをアレイグループ内のあるディスクドライブ（以下パリティディスクドライブと称す）10d1～10d3などに記録し、各アレイグループ内のあるディスクドライブ（故障ディスクドライブ）に故障が発生したとき、パリティディスクドライブ10d1～10d3などからパリティデータを読み出し、そのデータを基に故障ディスクドライブのデータを復元し、スペアディスク10e1～10e3などに書き込むための制御を行う。6はバッファ制御回路であり、バッファメモリを制御するものである。7はデータバッファメモリであり、ディスクドライブから読み出したデータを一時的に格納するバッファメモリである。8はディスクインタフェース回路（以下インタフェースをI/Fと称す）であり、ディスクドライブ側のI/Fを制御するものである。9はホストI/F回路であり、ホスト装置1側のI/Fを制御するものである。10a1～10e3はディスクドライブであり、ディスクコントローラ2によりSCSII/F12を通じてビットデータがリード/ライトされる。なおディスクドライブ10d1～10d3には、それぞれのグループのパリティデータが記録される。11はホスト装置側のSCSII/Fであり、ホスト装置1とのI/Fである。12はディスクドライブ側のSCSII/Fであり、1つのアレイグループ内のディスクドライブ台数分、ディスクコントローラ2に設けられている。13a～13eは交換単位となる複数のユニットであり、それぞれが所定数のディスクドライブ、例えば3台のディスクドライブ10#1～10#3などを収容して構成されている。交換の単位となるのは縦の列単位（1ユニット）である。14a～14cはアレイグループであり、

交換単位とは異なる列、つまり横の列で決められており、RAID機能によりパリティを生成するディスクドライブの論理的なグループである。15は内部システムバスであり、ディスクコントローラ2内のデータバスである。

【0029】ここで、RAID機能について説明する。

【0030】一般に、RAID機能のレベルとしては、レベル1 (RAID1) からレベル5 (RAID5) などがある。

【0031】RAID1は、従来より広く採用されているミラーリング (2重化ディスク、ミラードディスクなどと称するもの) である。

【0032】RAID2は、メモリにてしばしば使用されるECCの機能をディスクアレイに適用したものである。複数台のデータHDD (ハードディスクドライブ) に対し、ECCチェックデータとなる複数台のチェックHDDを冗長付加し、ECC機能によりデータを復元する。但し、このレベルは、一般的に実際のソリューションとしては考えられておらず、現在、実現製品はほとんど存在していない。

【0033】RAID3は、ユーザデータをストライピングにより複数台のHDDに分散して格納する。またこれらHDDグループに対しパリティデータ専用のHDDを付加し、データを復元する機能である。ストライピングサイズは、1バイトもしくはブロックであり、ユーザデータが複数台のHDDに対してパラレルに格納されると同時にパリティを生成し、これをパリティHDDに格納する。この機能のメリットは、HDDに対してパラレルにアクセスするので高速なデータ転送レートが実現できる。

【0034】RAID4は、RAID3と同様にストライピング+パリティ専用のHDDを有するが、そのストライピングサイズが比較的大きく設定された機能である。1回のユーザアクセスは、RAID3がアレイ内のHDDへパラレルにアクセスされるのに対してRAID4では、ストライピングサイズが大きいのでアレイ内の一部のHDDのみにアクセスされる。この機能のメリットは1回のユーザデータアクセスがアレイ内の一部のHDDに対してのみ実施されるため、残りのHDDへのユーザアクセスが可能となり、複数リクエストの同時実行が可能となる。このレベルは一般的に実際のソリューションとしては考えられておらず、現在、実現製品はほとんど存在していない。

【0035】RAID5は、RAID3と同様のストライピングサイズによるストライピングを有するが、パリティデータを格納するHDDが固定されずに、アレイ内の各HDDにパリティデータを分散して格納する機能である。この機能のメリットは、RAID4がパリティHDDが固定されているため、そのパリティHDDへのアクセスが性能的なボトルネックとなるのに対し、RAI

D5では、パリティデータが各HDDへ分散されて格納されるため、特定のHDDがボトルネックとなることはない。

【0036】次に図2を参照してこのディスクアレイサブシステムの動作を説明する。

【0037】図2に示すように、このディスクアレイサブシステムでは、ある1台のディスクドライブ、例えばディスクドライブ10c3などが故障した場合、ディスクコントローラ2は、故障したディスクドライブ10c3と同一のアレイグループ14c (パリティを生成する対象となるディスクドライブのグループ) に属する故障したディスクドライブ10c3以外のディスクドライブ10a3、10b3、10d3の内容 (“1”、“0”などのビットデータ) を読み出す。そしてパリティ制御回路5にて、ディスクドライブ10a3、10b3のビットデータの排他的論理和を取り、ディスクドライブ10d3のパリティデータと比較することにより、故障したドライブの内容 (データ) を復元し、そのデータをバッファ制御回路6がデータバッファメモリ7に書き込む。

【0038】続いて、ディスクコントローラ2は、このデータバッファメモリ7に書き込まれたデータを、故障ディスクドライブ10c3に対応するスペアユニット13e内のディスクドライブ (スペアディスクドライブ) 10e3に書き込む。

【0039】次にディスクコントローラ2は、故障ディスクドライブ10c3と同一のユニット13cの他のディスクドライブ10c1、10c2の内容をそれぞれ対応するスペアユニット13e内のディスクドライブ10e1、10e2へコピー (複写) する。なお上記した復元処理ならびにコピー処理などは、ホスト装置1からのアクセス要求がないときにディスクコントローラ2が実行する。

【0040】これらの復元処理と複写処理が完了すると、故障ディスクドライブ10c3を含むユニット13cの機能は、スペアユニット13eにより完全に代替されるようになり故障ディスクのユニット13cは使用されなくなる。

【0041】したがって、故障ディスクドライブ10c3を新しいものに交換するために、故障ディスクドライブ10c3を収容したユニット13cを、SCSI I/F 12から抜去した場合でも、スペアユニット13eによりその機能が完全に代替されているので、ディスクドライブ10c3が故障する以前の状態と同様にシステム全体が動作し、性能に影響なくシステムを運用することができ

る。

【0042】ここで、もし従来の制御形態のまま、このように複数のディスクドライブを収容したユニットで交換を実施すると、故障ディスクドライブのデータはスペアのディスクドライブによって代替されるものの、その他のディスクドライブについてはそのまま使用されているため、ユニット抜去後に、スペアに複写されなかった

ディスクドライブ10c1やディスクドライブ10c2に対するデータの復元処理が実施されることになり、この復元処理のためにディスクシステムそのものの性能低下が発生する。

【0043】一方、本実施例では、同一I/Fに接続された複数のディスクドライブを1ユニット化し、システムが稼働状態でも性能および信頼性に影響を与えることなく、複数のディスクドライブをユニットごとの交換可能にし、これによりキャビネットの奥行き方向の空間がディスクドライブの実装用の空間として利用できるようになり、多くのディスクドライブがキャビネット内に実装できるようになる。

【0044】また故障ユニット13cをSCSII/F12から抜去したときにノイズが発生しても、同一I/Fに接続されている全てのディスクドライブ10c1～10c3と一緒に取り外してしまうので、そのときのノイズは他に影響を及ぼさなくなり性能劣化が起こることもない。

【0045】このように本実施例のディスクアレイサブシステムによれば、2台以上のディスクドライブ10a1～10a3をSCSII/F12に複数接続して1ユニット化し、そのユニットをアレイ状に複数配置し、複数のユニット13a～13e中にスペアユニット13eを設定し、複数の中のあるディスクドライブ、例えばユニット13cのディスクドライブ10c3などが故障したとき、故障したディスクドライブ10c3のデータを復元して対応するスペア用のディスクドライブ10e3に書き込み後、故障ディスクドライブ10c3を格納したユニット13c内の他のディスクドライブ10c1、10c2のデータをスペアユニット13eのそれぞれのディスクドライブ10e1、10e2にコピーし、スペアユニット13eを代替動作させることにより、システム稼働中でも性能に影響することなく故障ディスクドライブ10c3をユニット13cごと交換することが可能となる。

【0046】したがって、図3に示すように、複数のディスクドライブ10a1～10a3をキャビネット20内の奥行き方向に収容したユニット13aをアレイ状に複数列設できるようになり、図4に示した従来のキャビネット40内で無駄な空間となっていた奥行き方向のスペース45をディスクドライブの実装スペースとして有効に使用することができる。

【0047】これによりディスクドライブシステムの実装密度が向上すると共に、装置形状の制約が軽減されるなどの効果が得られる。

【0048】なお上記実施例では、RAID機能レベル3またはレベル4のパリティ構成について説明したが、本発明がその他のレベルについても上記同様に適用できることは言うまでもない。

【0049】

【発明の効果】以上説明したように請求項1記載の発明によれば、ディスクドライブはディスクコントローラに

対してグループ間の列毎にまとめて着脱自在にユニット化されているので、ディスクコントローラからユニットを取り外したときにノイズが発生しても他のディスクドライブへの影響がなくなり性能劣化は起こらなくなる。また請求項2記載の発明によれば、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後にディスクコントローラが複数の中の所定ユニットをスペアユニットとして運用するので、不具合が発生したユニットを抜去しても、システム全体の性能が低下することがなくなる。

【0050】さらに請求項3記載の発明によれば、復元されたデータと、故障ディスクドライブの属するユニット内の他のディスクドライブのデータとがデータ書込手段により複数の中の所定ユニットのそれぞれ対応するディスクドライブに書き込まれて、所定ユニットが故障ディスクドライブの属するユニットの代替えとして運用されるので、故障ディスクドライブの属するユニットごと交換しても、性能劣化が起こらなくなる。

【0051】また請求項4記載の発明によれば、復元されたデータは、まず、第1の書込手段によりスペアユニットの対応するディスクドライブに書き込まれ、続いて第2の書込手段により、故障ディスクドライブが属するユニット内の他のディスクドライブのデータがスペアユニットのそれぞれ対応するディスクドライブに書き込まれて、そのスペアユニットが故障ディスクドライブの属するユニットの代替えとしてシステム上で運用されるので、故障ディスクドライブの属するユニットごと交換しても、性能劣化が起こらなくなる。

【0052】上記により、故障ディスクドライブを新しいものに交換するためにユニット単位で取り外してもノイズの影響がなくなり、システム動作中でも他のユニットに影響を与えることなく交換することができる。

【0053】この結果、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することができる。

【図面の簡単な説明】

【図1】本発明に係る一実施例のディスクアレイサブシステムの構成を示す図である。

【図2】このディスクアレイサブシステムの動作を示す図である。

【図3】このディスクアレイサブシステムにおいて、筐体内に複数のディスクドライブを実装した例を示す図である。

【図4】従来のディスクアレイサブシステムの実装例を示す図である。

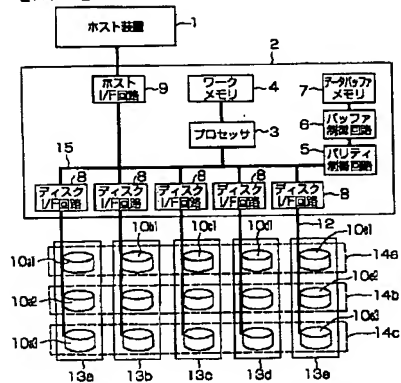
【符号の説明】

1…ホスト装置、2…ディスクコントローラ、3…プロセッサ、4…ワークメモリ、5…パリティ制御回路、6…バッファ制御回路、7…データバッファメモリ、8…ディスクI/F回路、9…ホストI/F回路、10…ディス

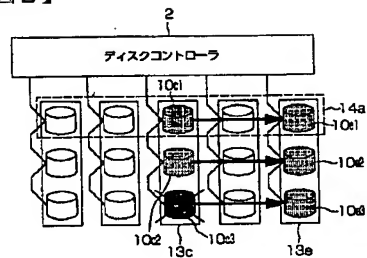
クドライブ、11、12…SCSI I/F、13a~13e…交換単位ユニット、14a~14c…アレイグルー

プ。

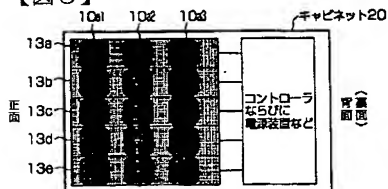
【図1】



【図2】



【図3】



【図4】

